

Roles of Adeno-Associated Virus Rep Protein and Human Chromosome 19 in Site-Specific Recombination

SAMUEL M. YOUNG, JR.,^{1,2} DOUGLAS M. McCARTY,² NATALYA DEGTYAREVA,³
AND RICHARD JUDE SAMULSKI^{1,2,3,4*}

*Curriculum in Genetics and Molecular Biology,¹ Lineberger Comprehensive Cancer Center,³
Department of Pharmacology,⁴ and the Gene Therapy Center,² University of
North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599*

Received 13 December 1999/Accepted 20 January 2000

Adeno-associated virus type 2 (AAV) is the only known eucaryotic virus capable of targeted integration in human cells. AAV integrates preferentially into human chromosome (ch) 19q13.3qter. The nonstructural proteins of AAV-2, Rep78 and Rep68, are essential for targeted integration. Rep78 and Rep68 are multifunctional proteins with diverse biochemical activities, including site-specific binding to AAV and ch-19 target sequences, helicase activity, and strand-specific, site-specific endonuclease activities. Both a Rep DNA binding element (RBE) and a nicking site essential for AAV replication present within the viral terminal repeats are also located on ch-19. Recently, identical RBE sequences have been identified at other locations in the human genome. This fact raises numerous questions concerning AAV targeted integration; specifically, how many RBE sequences are in the human genome? How does Rep discriminate between these and the ch-19 RBE sequence? Does Rep interact with all sites and, if so, how is targeted integration within a fixed time frame facilitated? To better characterize the role of Rep in targeted integration, we established a Rep-dependent filter DNA binding assay using a highly purified Rep-68 fusion protein. Electron microscopy (EM) analysis was also performed to determine the characteristics of the Rep-RBE interaction. Our results determined that the Rep affinity for ch-19 is not distinct compared to other RBEs in the human genome when utilizing naked DNA. In fact, a minimum-binding site (GAGYGAGC) efficiently associated with Rep, suggesting that as many as 2×10^5 sites may exist. In addition, such sites also exist frequently in nonprimate mammalian genomes, although AAV integrates site specifically into primate genomes. EM analysis demonstrated that only one Rep-DNA complex was formed on ch-19 target DNA. Surprisingly, identically sized complexes were observed on all substrates containing a RBE sequence, but never on DNA lacking an RBE. Rep-DNA complexes involved a multimeric protein structure that spanned ca. 60 bp. Immunoprecipitation of AAV latently infected cells determined that 1,000 to 4,000 copies of Rep78 and Rep68 protein are expressed per cell. Comparison of the Rep association constant with those of established DNA binding proteins indicates that sufficient molecules of Rep are present to interact with all potential RBE sites. Moreover, Rep expression in the absence of AAV *cis*-acting substrate resulted in Rep-dependent amplification and rearrangement of the target sequence in ch-19. This result suggests that this locus is a hot spot for Rep-dependent recombination. Finally, we engineered mice to carry a single 2.7-kb human ch-19 insertion containing the AAV ch-19 target locus. Using cells derived from these mice, we demonstrated that this sequence was sufficient for site-specific recombination after infection with transducing vectors expressing Rep. This result indicates that any host factors required for targeting are conserved between human and mouse. Furthermore, the human ch-19 *cis* sequences and chromatin structure required for site-specific recombination are contained within this fragment. Overall, these results indicate that the specificity of targeted recombination to human ch-19 is not dictated by differential Rep affinities for RBE sites. Instead, specificity is likely dictated by human ch-19 sequences that serve as a Rep protein-mediated origin of replication, thus facilitating viral targeting through Rep-Rep interactions and host enzymes, resulting in site-specific recombination. Control of specificity is clearly dictated by the ch-19 sequences, since transfer of these sequences into the mouse genome are sufficient to achieve Rep-dependent site-specific integration.

Adeno-associated virus type 2 (AAV) contains a single-stranded DNA genome of approximately 4.7 kb (50) and is a member of the *Parvoviridae* family (3). AAV is unique among other eucaryotic DNA viruses in that it utilizes a biphasic lifecycle to persist in nature. In the presence of a helper virus, adenovirus (Ad) or herpesvirus, AAV will undergo a productive infection. In the absence of a helper virus, AAV will integrate preferentially (>70%) into chromosome (ch)

19q13.3qter (3, 35). The ability of this nonpathogenic DNA virus, or virus-derived vector systems, to integrate site specifically have made it an attractive candidate vector for human gene therapy (45).

The AAV genome consists of two open reading frames (ORFs), which comprise the *rep* and *cap* genes, and 145-bp inverted terminal repeats (ITRs), which serve as the origins of replication (3, 35). The left ORF of AAV encodes four nonstructural proteins, Rep78, Rep68, Rep52, and Rep40. Extensive characterization of Rep78 and Rep68 *in vitro* has identified the following biochemical activities, DNA binding (18, 19), site-specific and strand-specific endonuclease activities (17, 19), and DNA-RNA and DNA-DNA helicase activities (17, 19, 59), all of which appear to be necessary for viral replication

* Corresponding author. Mailing address: Gene Therapy Center, 7119 Thurston Bowles CB 7352, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599. Phone: (919) 962-3285. Fax: (919) 966-0907. E-mail: rjs@med.unc.edu.

(15, 53). More importantly, Rep78 and Rep68 are required for mediating targeted integration (2, 43, 47, 51, 60).

Though site-specific integration is dependent upon either of the two large Rep proteins, the AAV ITRs are the only *cis* elements required for integration (34, 44, 61). In the absence of Rep proteins, the virus will still integrate through the ITR sequence but randomly into the host genome (21, 56, 61). Although integration in the absence of the Rep proteins is random, virus-cell junctions are nearly identical to junctions formed during targeted integration (DNA microhomology at junctions, specific deletions of the ITR sequences, rearrangement of the chromosome locus, and head-to-tail virus concatamers) (41, 62). In fact, *in vitro* integration products generated using cellular extracts produced identical type junctions, demonstrating the essential role the ITRs play in viral integration (62). From this analysis, Yang et al. (62) concluded that both random and targeted integration are dependent upon a cellular recombination pathway, with the role of Rep facilitating integration at ch-19. To help account for AAV targeting, a nearly identical Rep binding element (RBE) and a nicking site (*trs*) to that present on the AAV ITR was identified on the ch19.13.3qter AAV integration sequence (23–25, 43, 46, 54, 57). It was also demonstrated that Rep68 could mediate complex formation between the AAV ITR and the ch-19 integration site *in vitro* (57). This led to a hypothesis that AAV may target integration by Rep-mediated complex formation between the AAV ITR and the ch19 integration site. However, since this observation subsequent data has demonstrated that Rep can bind to degenerate RBE sequences, (5, 32). In fact, computer analysis identified at least 15 genomic genes which contained RBE sites that bound to AAV Rep protein *in vitro*, all more efficient than the ch-19 sequence (58). These data raise the question as to how Rep can target ch-19 among other RBE sequences. Using an Epstein-Barr virus (EBV)-based shuttle vector system carrying sequences from ch-19, Linden et al. demonstrated that the *trs* site was also critical for AAV site-specific integration (29, 30). When the *trs* site was not present, targeting was lost, even though the RBE was present. The present study suggested that both sequences were essential for site-specific integration (the RBE and the *trs* sequences). The probability of identifying a RBE with the correct proximity of a *trs* site would suggest a frequency of $<6 \times 10^{-11}$ /genome, thereby defining a unique sequence in the human genome (54). While these studies identify ch-19 *cis* elements required for AAV targeted integration and suggest why this reaction is specific, how Rep carries out this reaction remains unclear.

Critical to any model of AAV Rep-mediated targeted integration is the ability to recognize the ch-19 target sequence among other potential RBE sequences. Though Rep can bind many degenerate sequences, the actual definition of what constitutes an RBE is somewhat unclear. Random oligonucleotide selection demonstrated that the RBE could be defined as an 8-bp sequence: 5'-GAGYGAGC-3' (5). However, it was shown by methylation interference assays that the RBE was an 18-bp core sequence and that any mutation within this sequence would significantly affect Rep binding (42). Also, the report by Wonderling and Owens (58) demonstrated that the RBE oligonucleotides derived from the BLAST search contained mutations in this 18-bp core sequence but still bound better to the MBP-Rep68 than to the ch-19 RBE. Depending on the definition of an AAV RBE, the copy number present in the human genome (GAGYGAGC = 200,000 copies/genome, whereas 18-bp core = 1 copy/genome) could significantly impact the ability of Rep to identify its target locus.

Based on the above information, the number of RBE se-

quences in the human genome, how Rep discriminates between these and the ch-19 target locus RBE sequence, and how Rep interacts with all sites and still facilitates targeted integration within a fixed time frame become of significant importance. In this study, we evaluated the role of alternative RBEs in the human genome and how these sequences might impact the ability of Rep to target the locus on ch-19. Using a filter-binding assay and a highly purified source of Rep68 protein, we established that genomic DNA will compete efficiently against a ch-19 target sequences. In this assay, a minimum Rep binding site of 8-bp in the context of large DNA fragments demonstrated competition, suggesting that as many as 200,000 potential binding sites may exist in the human genome. Filter-binding analysis of genomic DNA successfully retained ch-19 target sequences, as well as a cellular RBE identified by BLAST analysis, corroborating the competition results. Electron microscopy (EM) analysis was utilized to distinguish possible differences between Rep protein DNA interaction with ch-19 RBE compared to a minimum 8-bp RBE sequence. Identical multimeric Rep protein DNA complexes, which spanned about 60 bp, assembled on ch-19 target DNA, as well as a minimum RBE site, but never on heterologous DNA lacking these sequences. At a high Rep concentration, protein DNA looping structures were detected, but no evidence for paranemic structures were observed. *In vivo* analysis of Rep protein levels in a latent infection demonstrated approximately 1 to 4,000 copies/cell. Analysis of Rep expression in non-virus-infected cells demonstrated DNA rearrangement of the ch-19 target sequence, suggesting that this locus is a hot spot for Rep-induced DNA amplification and rearrangement that most likely influences AAV targeted integration. Finally, generation of an animal model carrying the human ch-19 sequence at the mouse hypoxanthine phosphoribosyltransferase (HPRT) locus facilitated AAV Rep-mediated targeted integration and corroborates the importance of the ch-19 RBE-*trs* sequence.

MATERIALS AND METHODS

Plasmids. The pRE2 plasmid contains the 2.66-kb *Bam*HI subcloned fragment from chromosome 19 (63). pAB11 plasmid (10) contains the 3.1-kb *lacZ* gene. The pcSF17 plasmid (ATCC 53149) contains the cDNA of CSF1. The pT7-7BT plasmid is a derivative of the pT7-7 vector (1) but has the degenerate RBE removed from its plasmid origin. SSV9 int- was constructed by replacing the *Xba*I fragment, which contained the AAV *rep* and *cap* genes with the *Xba*I fragment from plasmid dl-int- (a gift from N. Muzycka). In this mutant, the intron sequence has been specifically deleted. Thus, only Rep68 and Rep40 are expressed (60). Plasmids were digested with the appropriate enzymes to generate substrates and probes.

***Escherichia coli* expression vectors.** The AAV Rep68 coding sequences were cloned into the pQe70 vector (Qiagen). The AAV Rep68 ORF was PCR amplified from SSV9 int- using Vent polymerase (New England Biolabs). Oligonucleotides were 5'-ACCATGCATGCCGGGGTTTACGAG-3', which hybridizes to nucleotides 320 through 337 of SSV9 int-, and 5'-ACCATAGATCTGAGAGAGTGTCTCGAGC-3', which hybridizes to nucleotides 1910 through 1927 of SSV9 int-. The PCR product was digested with *Sph*I and *Bgl*II (NEB) and cloned into the pQe70 vector to generate a Rep68 histidine-tagged fusion protein, which introduced eight new amino acids (RSHHHHHH) at the carboxy end. This plasmid, pStump68, uses the original ATG of Rep68 and places the Rep68H6 expression under IPTG (isopropyl- β -D-thiogalactopyranoside)-inducible control. pStump68 (Fig. 1) was transfected into SG13009 (13) (a gift of S. Gottesman), which contained the pRep4 plasmid (Qiagen). pStump68 was also sequenced and was determined to contain a mutation that resulted in an S536A mutation (University of North Carolina [UNC] sequencing facility). Expression of full-length fusion protein was confirmed by immunoblotting with Rep-specific monoclonal antibodies (16).

Rep-specific antibodies. The IF11 antibody that recognizes all four Rep proteins has been previously described (16). The Rep polyclonal antibody was generated by immunizing a rabbit three times with 150 μ g of the Rep68H6 protein each time (Spring Valley Laboratories).

Preparation of *E. coli* extracts. One liter of Stump68 cells, SG13009 cells that contain the pRep4 and pStump68 plasmids, were grown at 37°C to an optical density of 0.8 and were induced for 1 h with 0.1 mM IPTG. Cells were washed in 50 mM NaPO₄ (pH 8.1)–1 M NaCl and pelleted again. Cells were then

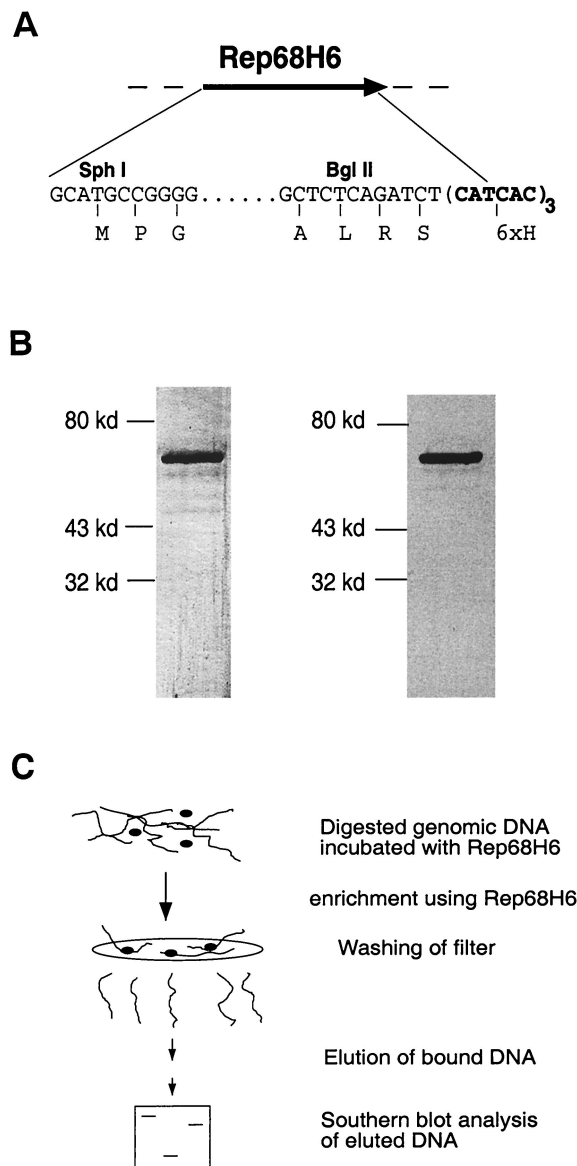


FIG. 1. Filter-binding assay. (A) Diagram of the modification to the AAV Rep68 protein in the expression plasmid pStump68. pStump68 is a fusion construct carrying a His tag (six amino acids) placed at the C terminus. The relevant restriction sites (*Sph*I and *Bgl*II), 5' and 3' DNA sequence, and amino acids (letters below the DNA sequence) utilized in the cloning approach are illustrated. See Materials and Methods for details concerning the construction. (B) Silver staining and Western blot analysis of purified Rep68H6. The left part of panel B is a silver stain of 250 ng of purified Rep68H6. By silver stain analysis, Rep68H6 was purified to apparent homogeneity. The right panel is a Western blot showing that the band that appears on the silver stain is Rep specific. More details concerning the purification can be found in the Materials and Methods. (C) Rep filter-binding assay for RBE containing genomic fragments. Essentially, purified Rep68H6 is incubated with digested HeLa genomic DNA (details are in Materials and Methods), and the protein-DNA reaction is then passed over a nitrocellulose filter and washed with a nonelution buffer. Rep-specific bound DNA is eluted and fractionated on agarose gel and then analyzed by Southern blotting.

resuspended in 50 mM NaPO₄-1 M NaCl-0.1% Tween 20-10 mM β -mercaptoethanol (BME)-50 mM imidazole (pH 7.0)-0.5 μ g of leupeptin per ml-0.7 μ g of pepstatin A per ml-0.1 mM phenylmethylsulfonyl fluoride (PMSF). Cells were then subjected to two freeze-thaws and incubated on ice in the presence of 1 mg of lysozyme per ml for 30 min. They were then sonicated for 30 s on ice at an output of 6 and a duty cycle of 50 on a Branson Sonifier 250 (VWR Scientific). The lysate was centrifuged at 25,000 rpm in a Beckman SW41 centrifuge at 4°C

for 30 min. Glycerol was added to a final concentration of 20% to the soluble cell lysate, which was then frozen at -80°C or used immediately.

Purification of Rep68. The soluble *E. coli* extract was applied to an Ni²⁺-nitriloacetic acid (NTA) (Qiagen) column (bed volume, 2 ml; diameter, 1.0 cm). The column was equilibrated in 50 mM NaPO₄-1 M NaCl-0.1% Tween 20-10 mM BME-50 mM imidazole (pH 7.0)-20% glycerol (equilibration buffer) or, in later experiments, the buffer had a final pH of 8.1. The *E. coli* extract was then applied to the column at a flow rate of 0.20 ml/min. The column was washed with 5 column volumes of mixture containing 50 mM NaPO₄, 1 M NaCl, 0.1% Tween 20, 10 mM BME, 100 mM imidazole (pH 7.0), 20% glycerol, 0.5 μ g of leupeptin per ml, 0.7 μ g of pepstatin A per ml, and 0.1 mM PMSF; the buffer was then adjusted to a final pH of 6.0 (wash buffer) at flow rate of 0.5 ml/min. The protein was eluted at a flow rate of 0.5 ml/min in an ascending linear gradient of 0.1 to 1 M imidazole in the wash buffer, except that the final pH was adjusted to 8.1. The Ni-NTA fractions, which contained Rep68H6, were identified by silver staining (Bio-Rad Silver Stain Plus). For some experiments the Rep68H6 fractions were then pooled, dialyzed into 25 mM Tris-HCl (pH 7.5)-50 mM NaCl-10 mM BME-20% glycerol-0.5 μ g of leupeptin per ml-0.7 μ g of pepstatin A per ml-0.1 mM PMSF, frozen at -80°C and designated Rep68H6 Nickel.

Rep68H6 Nickel fractions were dialyzed into 25 mM Tris-HCl (pH 8.0)-100 mM NaCl-0.1% Tween 20-0.1 mM EDTA-10 mM BME-20% glycerol (final pH, 8.1) at 4°C (MonoQ equilibration buffer). The equilibrated Rep68 was then applied to a 1-ml MonoQ fast-protein liquid chromatography (FPLC) column (Pharmacia). The dialyzed Rep68H6 was applied to the MonoQ column at a flow rate of 0.5 ml/min. The column was washed with 5 column volumes of MonoQ equilibration buffer at flow rate of 1.0 ml/min. The protein was eluted in a linear gradient ascending from 0.1 to 1.0 M NaCl in MonoQ equilibration buffer. The contents of the peak were determined by silver staining (Bio-Rad Silver Stain Plus). The Rep68H6 peak was pooled and dialyzed into 25 mM Tris-HCl (pH 8.0)-400 mM NaCl-1 mM dithiothreitol-0.1% Tween 20-20% glycerol and stored at -80°C; this fraction was called Rep68H6. Both Rep68H6 Nickel and Rep68H6 were shown to possess wild-type activities as determined by DNA binding and *trs* endonuclease and DNA helicase assays (17). To determine Rep protein concentrations the bicinchoninic acid assay was employed (Pierce, Inc.).

Filter-binding assays. Filter-binding assays were performed as described by Fuller and his colleagues (11), with some modifications. The standard reaction conditions contained 20 pmol of Rep68H6 incubated in a 300- μ l reaction volume which contained 10 μ g of *Bam*HI-digested HeLa genomic DNA and 10 mM HEPES-NaOH (pH 7.8)-200 mM KCl-10% glycerol-10 mM BME. The reactions were filtered under gentle suction (0.4 ml/min) on presoaked filters (Millipore Type HA, 0.45 μ m [pore size]). Filters were washed with 10 ml of non-elution buffer (25 mM Tris-HCl, pH 8.0; 200 mM KCl; 10% glycerol) under gentle suction. The bound DNA was then eluted as previously described (8). Exceptions to these conditions are noted. The eluted DNA was then analyzed on agarose gels. In some experiments, various amounts of poly(dI-dC) (Pharmacia) or digested HeLa genomic DNA was added to the reaction. In other experiments, the concentration of KCl was varied. In some experiments various amounts of ³²P-end-labeled pRE2 fragment were included, while in other experiments the volume of wash buffer was varied. Finally, competition assays were performed using the 3.1-kb *lacZ* gene and the 2.7-kb ch-19 integration region, and the 938-bp colony-stimulating factor (CSF) cDNA were purified away from the plasmid backbone by digestion with appropriate restriction enzymes and isolated by gel purification using the Gene Clean Kit (Bio 101).

Southern hybridization analysis. Filter-binding assays, which were carried out with no labeled fragment, were electrophoresed in a 1.0% agarose gel along with 10 μ g of HeLa genomic DNA, which was not subjected to the filter-binding assays. Retained DNA was transferred to Gene Screen Plus (New England Nuclear) as recommended by the manufacturer and hybridized at 65°C with randomly labeled probes (Boehringer Mannheim) specific to the ch-19 integration region, the neomycin resistance gene, or exon one of CSF1 (20, 26). In some other experiments mouse genomic DNA, which contained the human preintegration region from ch-19 and were infected with AAV-Neo, was digested with *Xba*I instead of *Bam*HI. The digested mouse genomic DNA was analyzed for integration by probing the blots with sequence-specific probes for ch-19 and the neomycin gene. Site-specific integration was defined as cohybridization of the two probes to the same band. Amplification of the ch-19 sequences in the presence of Rep was determined by image quantification of bands compared to control (non-Rep) lanes. Any lane in which the sum of the intensity of the signals surpassed the intensity of the control lane was identified as amplified.

Preparation of protein-DNA complexes for EM. EM analysis was carried out at the UNC EM Core Facility. Binding reactions were carried out as previously described, except that reactions were 50 μ l and included the *Bam*HI-*Sca*I-digested pRE2 plasmid at a concentration of 1 μ g/ml and AAV no-end DNA substrates (49). The *Bam*HI-*Sca*I digestion released three fragments. The 2.66-kb ch-19 integration fragment, a 1.7-kb fragment which contains an RBE of 5'-GAGTGAGC-3', and a 953-bp nonspecific fragment. Rep protein was added to the reaction at a concentration of 25 monomers to 1 RBE.

Cross-linking of protein-DNA complexes was carried out by addition of glutaraldehyde to 0.6% and allowing the reaction mixture to incubate for 5 min at room temperature. After 5 min reactions were applied to Bio-Gel A-5m (Bio-Rad) columns, which had been equilibrated with 10 mM Tris-HCl (pH 7.6)-0.1 mM EDTA. The filtered samples were then incubated with a buffer containing 2

TABLE 1. Summary of Rep protein and DNA characteristics as determined by EM^a

Binding group (fragment size)	Protein/DNA mass ratio	Binding characteristics		
		% Specific binding	% Nonspecific binding	% Free DNA
2.7 kb	1:1	40	16	44
1.7 kb	1:1	48.4	11	40.6
953 bp	1:1	0	0	100

^a The 2.66-kb ch-19 integration contains an RBE located directly in the center of the fragment, while the 1.7-kb fragment contains an RBE located 180 nucleotides from one end of the fragment; the 953-bp fragment does not contain any known RBE. "Specific" was distinguished from "nonspecific" as any protein complex with the correct location on the 2.66-kb fragment. The same approach to analysis was used for the 1.7-kb fragment. Since the 953-bp fragment does not contain any known RBE, all Rep complex bound to this fragment was defined as nonspecific binding. A total of 100 DNA molecules were counted for each fragment. Binding was reflected as a percentage of the number of DNA molecules that were bound by Rep.

mM spermidine, 2 mM MgCl₂, 0.15 M NaCl, and 0.10 M KCl (28) and adsorbed to glow-charged thin carbon foils, dehydrated through a water-ethanol series and rotary shadow cast with tungsten as described elsewhere (14). Samples were visualized in a Philips CM12 instrument. Micrographs for publication were scanned from negatives by using a Nikon multifilm scanner; the contrast was optimized and the panels were arranged using Adobe Photoshop.

The 2.66-kb ch-19 integration contains an RBE located directly in the center of the fragment; the 1.7-kb fragment contains an RBE located 180 nucleotides from one end of the fragment, while the 953-bp fragment does not contain any known RBE. Specific was distinguished from nonspecific by any protein complex with the correct location on the 2.66-kb fragment. The same approach to analysis was used for the 1.7-kb fragment. Since the 953-bp fragment does not contain any known RBE, all of the Rep complex binding to this fragment was defined as nonspecific binding. A total of 100 DNA molecules were counted for each fragment. Binding was reflected as a percentage of the number of DNA molecules that were bound by Rep (Table 1).

Detection of Rep in latently infected cells. HeLa cells which were grown to 80 to 90% confluency on 10-cm plates were infected with a multiplicity of infection (MOI) of 10 infectious units of wild-type (wt) AAV or with an MOI of 10 infectious units of wt AAV and an Ad type 5 MOI of 20. Mock-infected cells were also carried to serve as a negative control. Infections were allowed to proceed for 24 h, and then the supernatant was removed. Cells were rinsed twice with ice-cold, phosphate-buffered saline (PBS; UNC Tissue Culture Facility). Cells were then manually scraped and pelleted.

Cell pellets were resuspended in 200 µl of lysis buffer (150 mM NaCl; 50 mM Tris-HCl, pH 8.0; 5 mM EDTA; 1.0% Triton X-100; 1 mM PMSF). Cells were incubated on ice for 20 min and then vortexed three times for 10 s each time. Lysate was transferred to 1.5-ml Eppendorf tubes and spun at 4°C for 10 min at 12,000 rpm to pellet the cellular debris. Cleared lysate was transferred to fresh Eppendorf tubes, 20 µl of Protein A/G Plus-Agarose beads (Santa Cruz Biotechnology) was added, and 2 µl of a Rep polyclonal antibody was added. Reactions were incubated for 2 h at 4°C. After 2 h the reactions were centrifuged at 2,500 rpm for 5 min. The supernatant was aspirated. Then, 20 µl of 2× sodium dodecyl sulfate sample buffer (containing a 1:10 dilution of BME) was added. Pellets were then boiled and loaded onto sodium dodecyl sulfate–10% polyacrylamide gel electrophoresis gels. Only half of the immunoprecipitation reaction was loaded for the Ad-AAV-infected cells. Gels were transferred and immunoblotted as previously described (16). Briefly, they were transferred at 500 mA for 20 min using a semidry apparatus (Bio-Rad). The blots were incubated with the IF11 antibody for 1 h and then incubated with an anti-mouse horseradish peroxidase antibody (Amersham) for 1 h. Protein was detected by addition of the Super Signal Chemiluminescent Substrate (Pierce). Blots were exposed to Bio-Max MR film (Kodak).

ch-19 amplification detection. Superfect reagent (Qiagen) was used for transfection. Transfections were carried out according to the manufacturer's protocol. Cells were grown to 50 to 80% confluency. A total of 10 µg of DNA was used for transfection. Specifically, 5 µg of pHIV78 (12) and 5 µg of the AAV-Neo construct with nonfunctional ITRs (pNeo) at a 1:1 ratio were used. DNA was dissolved in TE (10 mM Tris-HCl, pH 7.5; 0.1 mM EDTA) and diluted into 300 µl of medium lacking serum, proteins, or antibiotics. Next, 40 µl of Superfect reagent was added to the DNA solution and mixed in by vortexing. Samples were allowed to incubate for 5 to 10 min at room temperature. While the samples were incubating, the plates to be transfected were washed with PBS. After incubation, 3 ml of Dulbecco modified Eagle medium (DMEM; Sigma) supplemented with 10% fetal bovine serum (FBS; Sigma) was added to each tube containing the transfection complexes and then mixed by pipetting. This mixture was then added to the plates. The cells were incubated with the transfection complexes for

2 h. Fresh medium was added, and cells were maintained for 48 h at 37°C and in 5% CO₂ in a tissue culture incubator. Cells were washed four times with 10 ml of PBS and then diluted to appropriate dilutions for single cell cloning.

G418 selection was carried out on cells that had been transfected with plasmids containing the neomycin resistance gene. Cells were placed under 600 µg of G418 per ml for approximately 8 days. After colonies were picked, cells were placed in 300 µg of G418 per ml to maintain selection of the neomycin-resistant cells.

To pick single cell clones, cell were first transfected with the neomycin-resistant gene and then given 48 h to express the resistance gene and to recover from the transfection. Cells were trypsinized and then plated at densities of 1:10, 1:50, and 1:100. Cells were placed under G418 selection at 600 µg of G418 per ml. After 8 days of selection, when the colonies had grown to 30 to 50 cells/colony, the cells were isolated.

To pick the colonies, plates were viewed under low magnification, and colonies that were well isolated from other colonies were circled on the plate. The medium was then removed from the plate, and the cells were washed gently with PBS. Working from the bottom of the plate to the top of the plate, a 10-µl Art Reach tip filled with 10 µl of trypsin-EDTA (Gibco BRL) was used to scratch and pull up the colonies of cells. Approximately 10 to 12 colonies were picked, and then the plate was washed again with PBS to prevent the plate from drying out. Picked colonies were placed in 96-well tissue culture plates containing DMEM with 10% FBS and allowed to grow for approximately 24 to 48 h. After this the cells were placed under selection with 300 µg of G418 per ml. After cells had reached confluency, they were passed into dishes with sequentially larger wells.

RESULTS

Rep filter-binding assay. In order to evaluate the role of Rep in targeting AAV to ch-19 sequences compared to alternative RBE sites, we established a filter-binding assay. The filter-binding assay was dependent upon a highly purified source of Rep68 protein that could be easily generated. A bacterial overexpression system that included an affinity tag was utilized to achieve this goal. Rep68 protein was cloned into the pQe70 vector (Qiagen) which fuses a His₆ tag to the carboxy end. We placed the His₆ tag at the carboxy end since genetic and biochemical characterization had shown that this portion of the protein can be deleted and still maintain all known biochemical activities (15, 31, 37). The fusion protein Rep68H6 is wt in sequence except for a serine-to-alanine change at amino acid 536, followed by an 8-amino-acid His tag (RSHHHHHH) after amino acid 537 (Fig. 1A).

Rep68H6 was overexpressed and purified by using a two-step column purification procedure: Ni-NTA resin (Qiagen), followed by MonoQ FPLC column (as described in the Materials and Methods). Silver stain analysis determined that Rep68H6 had been purified to >95% homogeneity (Fig. 1B), and immunoblotting confirmed the identity of the purified material as Rep specific (Fig. 1B). The purified Rep68H6 was assayed and determined to have all biochemical activities, including binding to the AAV ITR and *tr*s endonuclease and helicase activities as previously described (17) (data not shown). The yield of purified Rep68H6 was approximately 2 to 5 mg/liter of bacterial culture.

The availability of highly purified Rep protein allowed establishment of a Rep-dependent filter-binding assay. In essence, digested genomic DNA was incubated with purified Rep68H6 (Fig. 1C) and passed over a filter under conditions where only DNA bound to protein was retained. The bound DNA was subsequently eluted and analyzed by Southern blot. However, before analyzing Rep68H6 ability to bind specific sequences within the context of genomic DNA, reaction conditions for specific Rep68H6 binding were established. The objective was to establish conditions that would bind all potential RBE sequences in the human genome. Using this approach we determined that 20 pmol of Rep68H6 was the optimal amount of protein required to bind as many as 200,000 other potential RBE sites (this is a theoretical number de-

duced from an 8-bp minimum RBE site; see below) in the genome (data not shown).

Previous studies demonstrating Rep ability to bind degenerate RBE sites (5, 33, 58) strongly suggests that genomic DNA should compete efficiently in binding assays spiked with labeled ch-19 target sequences. To test Rep68H6's ability to bind the ch-19 target sequence, competition assays were performed in the presence of either genomic or nonspecific competitor DNA [poly(dI-dC)]. As little as 5 μ g of genomic DNA competed away >85% of labeled ch-19 fragment (1 fmol) (Fig. 2A). Maximum amounts of poly(dI-dC) (sixfold higher) reached only 80% of the level of competition observed with 5 μ g of genomic DNA (Fig. 2A). These results demonstrate and corroborate earlier studies (58) that genomic DNA contains many potential RBEs, and that these sites can compete for Rep binding.

In an effort to characterize specificity of the genomic versus ch-19 Rep interactions, assays were carried out with increasing amounts of wash buffer. In the absence of Rep, no genomic DNA was retained on the filter after 10 ml of washing (Fig. 2B). Similar assays carried out in the presence of Rep determined that 90% of ch-19-specific sequences and 25% of total genomic DNA was retained even after a 20-ml wash (as determined by the optical density at 260 nm) (Fig. 2B). We recently demonstrated that Rep-DNA complexes are sensitive to salt concentration (data not shown) (62a). We therefore determined the salt sensitivity of Rep-genomic DNA complexes and Rep-ch-19 complexes using either unlabeled and labeled substrates, respectively. As the molar amount of KCl was increased, there was a comparable reduction of the amount of both the ch-19 fragment and genomic DNA retained on the filter (Fig. 2C). This suggested that the genomic DNA had a Rep-binding characteristic similar to that of the ch-19 fragment and displayed Rep-binding sensitivity to salt (62a). These studies support a Rep-dependent interaction with genomic DNA that is similar to the characterized RBE on ch-19-specific DNA.

Comparison of Rep68H6 binding preference for RBEs in genomic DNA. The competition studies carried out above suggested that a portion of genomic DNA could compete for Rep binding. It was assumed that this interaction was related to non-ch-19 RBE sequences. To determine the specificity of the genomic sequences competing in the above reaction, filter-binding reactions using genomic DNA were analyzed by using either a ch-19 or a genomic RBE probe identified by a BLAST search. A previous report using RBE oligonucleotides derived from GenBank sequences demonstrated positive gel-shifted complexes when using a Rep68 maltose fusion protein (58). In one case, an RBE site identified in exon 1 of CSF 1 gene (CSF1) and identical to ch-19 DNA (58) bound with higher affinity. Since all 15 RBE oligonucleotides identified by BLAST analysis bound with higher affinity than a ch-19 oligonucleotide, we tested in the filter-binding assay for the ability of Rep to bind ch-19 and CSF1 RBE sites. In this assay, the high-affinity RBE site of CSF1 was compared to the ch-19 target sequence in the context of digested genomic DNA.

The filter-bound endogenous ch-19 genomic DNA fragment was visualized after Rep-binding reactions by using Southern blot analysis. This result allowed us to use this signal as a reference for positive Rep binding. As an additional control, genomic DNA was digested with *PvuII* to separate ch-19 RBE sequences from ch-19 3'-flanking DNA before submitting it to filter-binding analysis. By hybridizing with 3' ch-19-specific probes, we could determine Rep binding to non-RBE containing DNA from the adjacent chromosomal region. *PvuII* digestion generates a 4.2-kb genomic fragment of ch-19, which does

not contain the ch-19 RBE, and an 898-bp fragment, which contains the RBE (Fig. 3A). Filter-binding assays were carried out, and Rep-retained DNA was eluted and fractionated on agarose gels and probed with a right-half ch-19 sequence-specific probe (Fig. 3B). Southern analysis demonstrated that while the 4.2-kb *PvuII* fragment can be detected in the total genomic DNA, no signal was observed in the filter-bound lane even after extended overexposure (data not shown).

Unlike the ch-19 3'-flanking sequences, 5' genomic DNA carrying the ch-19 RBE sequences (*BamHI* 2.7-kb fragment) was retained and identified after Southern blot analysis (Fig. 3B). More importantly, CSF1 genomic DNA (*BamHI* 2.7-kb fragment) originally identified by BLAST analysis as positive for an RBE sequence was also retained after Rep filter binding (Fig. 3D). Filter-bound samples were image quantified after Southern blot analysis. Approximately 25% of ch-19 genomic DNA was retained by Rep (Fig. 3B, lane 2). In contrast, only one-third the amount of CSF1 genomic DNA was retained (Fig. 3D, lane 2). While the number of chromosomes carrying the CSF sequence are unknown in our aneuploid HeLa cells, we previously determined that three copies of chromosome 19 are present (46). Based on this information, we calculated a minimum 2.7-fold difference in Rep affinity for ch-19 sequences over CSF1 genomic DNA.

These studies demonstrated that in the context of genomic DNA both RBE sequences were recognized by Rep and retained. This would imply that all RBE sites in the human genome compete for Rep binding and that this step is not sufficient to determine ch-19 site-specific integration. Although the ch-19 target sequences bound more efficiently (2.7-fold) than the identical RBE element located in the CSF gene, this difference was not sufficient enough to explain AAV targeting frequency. The binding affinity observed was specific since we saw no retention of the 3' ch-19 genomic DNA devoid of RBE sequences. In addition, using genomic DNA to compare ch-19 to CSF1 RBE sites, we observed a higher affinity for ch-19 than that established in an *in vitro* reaction using oligonucleotides (58). These results suggest that other ch-19 flanking sequences within this fragment may influence the recognition of Rep for the ch-19 RBE site.

Direct visualization of Rep-DNA complexes by EM. Since Rep binding to the specific ch-19 fragment within genomic DNA appeared to be more efficient than to the CSF fragment and since this observation was contrary to published data generated using 54-bp oligonucleotides, we utilized EM to evaluate whether targeting may be facilitated by other DNA elements present on ch-19 (i.e., incomplete RBE sequences). To determine if multiple Rep-binding sites or additional activities, such as Rep formation of paranemic structures of the target DNA, occurred after Rep-DNA binding, Rep68H6 and ch-19 DNA were analyzed for size and number of protein-DNA complexes. A plasmid (pRE2) which contains the ch-19 region was digested to yield three fragments of 2.7 kb (which contain the genomic ch-19 sequence with the target RBE element located in the center), 1.7 kb (which contain a minimal RBE of 5'-GAGTGAGC-3' in the plasmid origin and is located approximately 200 bp away from one end), and 953 bp (devoid of any RBE sequences). Specific binding versus nonspecific binding was determined for all three fragments (see Materials and Methods for details). Rep68H6 was added at protein/DNA mass ratio of 1:1 (25 monomers of Rep to 1 RBE). The binding reactions were glutaraldehyde fixed, shadowed with tungsten, and visualized on EM (Fig. 4). At a mass ratio of 1:1, we found approximately 50% of the RBE-containing fragments bound to Rep. Rep complexes were predominantly localized to the center of the 2.7-kb fragment (Fig. 4A), asymmetrically localized

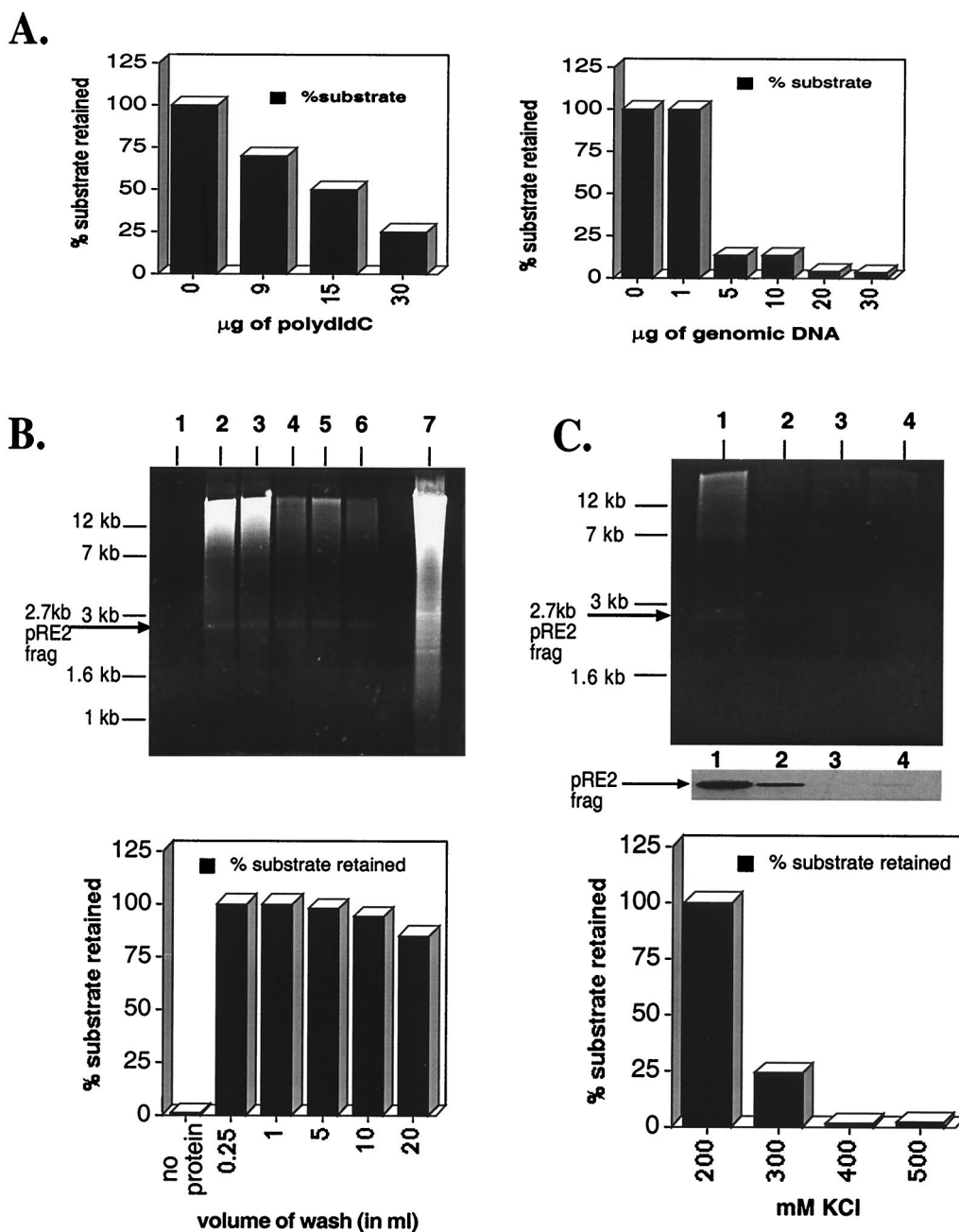


FIG. 2. Optimization of filter-binding reactions. (A) The left panel is a titration of increasing amounts of poly(dI-dC) required to disrupt Rep-ch-19 RBE-specific binding (as measured by filter binding). The reaction contains 20 pmol of Rep68H6 and 1 fmol of the labeled ch-19 target sequences. The right panel graphs the effect of increasing amounts of genomic DNA on Rep68H6-ch-19-labeled DNA as measured by the filter binding assay (see Materials and Methods for details). (B) Titration of nonelution buffer on Rep-ch-19-specific binding by ethidium bromide staining of agarose gel analysis (upper panel) and retention of radio labeled ch-19-specific probe (lower panel). Lane 1, no Rep68H6 protein; lanes 2 to 6, 0.25, 1.0, 5.0, 10, and 20 μg of nonelution wash buffer, respectively. Lane 7 shows 10 μg of *Bam*HI-digested HeLa genomic DNA prior to filter binding and elution. In all the reactions, 50 fmol of the ^{32}P -labeled ch-19 fragment from pRE2 was spiked in to serve as an internal control of Rep-specific binding. Results are graphed in the lower panel. (C) KCl effect on Rep genomic DNA filter-binding assay using ethidium bromide staining of agarose gel analysis (upper panel) and retention of radio labeled ch-19-specific probe (lower panel). Lane 1, 200 mM KCl; lane 2, 300 mM KCl; lane 3, 400 mM KCl; lane 4, 500 mM KCl. Reactions were carried out as described in panel A and in Materials and Methods and was followed by washing with 10 ml of nonelution buffer. The middle part of panel C shows an autoradiogram of retained ch-19 fragment. The lower panel shows a graph of the percentage of the ch-19 fragment retained with respect to increasing amounts of KCl.

on the 1.7-kb fragment (Fig. 4B), and never localized to the 953-bp fragment, as expected (Fig. 4C) ($n = 100$). Rep-DNA interaction at these ratios did not form gross structural modifications of the target DNAs (i.e., loop structures or intermolecular complexes). The size of the Rep complex bound to

DNA was deemed to be a multimer and spanned an average of 60 bp. This conclusion was based upon comparison to other known DNA-binding proteins bound to their respective targets and analyzed by EM, as well as by measuring the area of DNA covered by the protein complex (data not shown and J. Griffith,

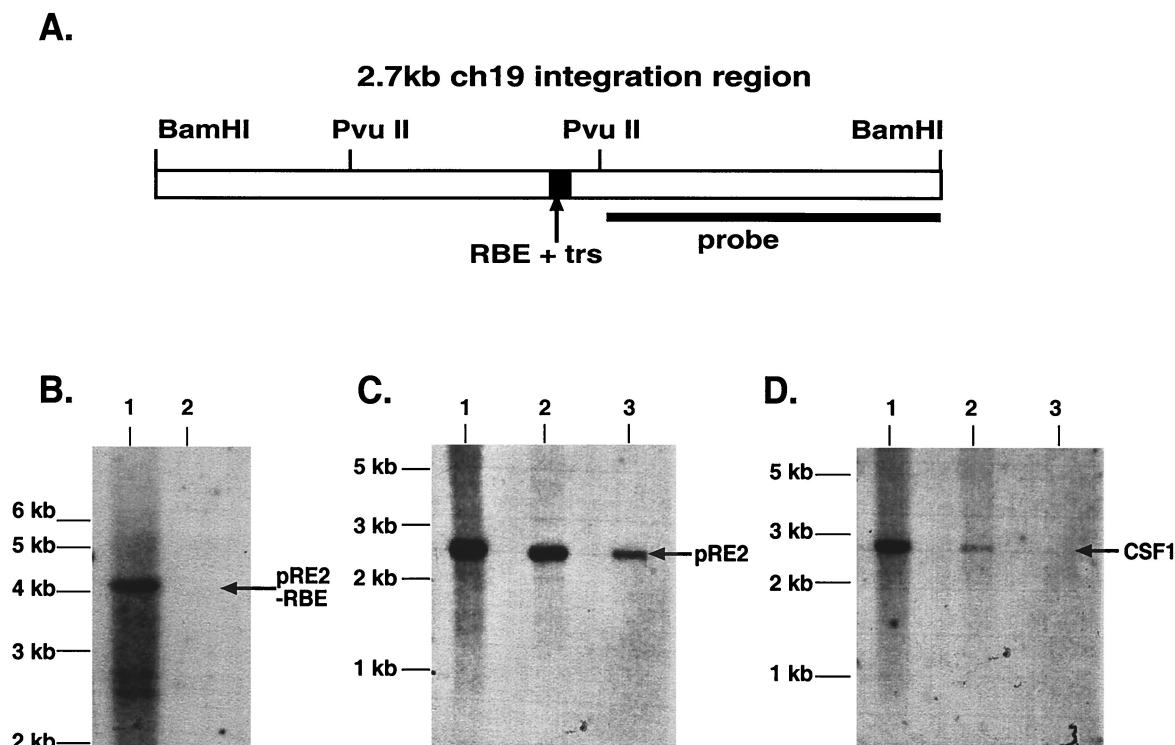


FIG. 3. Southern Blot analysis of filter-bound genomic DNA. (A) Diagram of the 2.7-kb ch-19 integration region. The RBE-*trs* is labeled and shown as a black box, while the sequence recognition region of the probe used for all analysis is depicted as a solid line. The distance between the two *PvuII* sites is 898 bp. (B) Filter-binding assay of ch-19 sequences devoid of any known RBE sequences. Genomic DNA digested with *PvuII* (one site in the 2.7-kb fragment and the other 3' to the *BamHI* site) is specific for a 4-kb ch-19 fragment devoid of RBE sequences. Lane 1, signal generated with 10 μ g of *PvuII*-digested HeLa genomic DNA prior to filter binding assay; lane 2, signal generated after loading DNA eluted from one filter-binding reaction. Hybridization was carried out with the *PvuII*-*BamHI* right-side-specific probe (see panel A). (C) Southern analysis demonstrating Rep68H6's ability to retain ch-19 sequences carrying RBE sequences after the filter-binding assay. Lane 1, signal generated from 10 μ g of *BamHI*-digested HeLa genomic DNA prior to filter binding; lane 2, signal generated from four pooled filter-binding reactions; lane 3, signal generated from one filter-binding reaction. The arrow indicates the position of the ch-19 fragment (pRE2). (D) Rep68H6 retention of non-ch-19 RBE genomic sequences. Lane 1, signal generated from 10 μ g of *BamHI*-digested HeLa genomic DNA prior to filter-binding reaction; lane 2, signal generated from four pooled filter-binding reactions; lane 3, signal generated from one filter-binding reaction. The blot was probed with a *SmaI*-*PstI* fragment from pcCSF17, which recognizes exon 1 from the CSF1 gene (see Materials and Methods for details). The arrow indicates the position of the CSF1 fragment.

personal communication). These results demonstrate that there are no other RBEs present on the ch-19 fragment, nor are there distal sequences (loop structures) stabilizing Rep-DNA interactions under these Rep concentrations. However, when Rep-AAV ITR sequences were analyzed using 1:1 ratios, we routinely observed that 100% of the substrates bound with the Rep complex to both of the terminal repeats (Fig. 5). In some of the AAV molecules, we also observed binding to a position that would correlate to the p5 promoter. The presence of these complexes and the appearance of DNA looping structures were observed (data not shown). As previously described, we saw a higher affinity of Rep for the AAV ITR sequences under these conditions than for the ch-19 and non-ch-19 RBE fragments. However, thorough EM analysis ($n = 100$) also demonstrated that there was no preference in Rep binding to the ch-19 fragment versus the 1.7-kb fragment containing the minimum 8-bp RBE (Table 1). This result is in contrast to the results of our Rep genomic filter-binding assay and may reflect the limited quantitative value of these protocols.

Minimum RBE sites recognized by Rep. By EM analysis, the 8-bp plasmid RBE bound Rep (Table 1) and generated protein-DNA complexes identical to ch-19 target DNA (Fig. 4). This result suggests that Rep is unable to discriminate between the ch-19 target RBE and other non-ch-19 sites. In addition, these results suggest that there are a multitude of potential RBEs in the human genome. Competition assays were per-

formed as described in Materials and Methods using a 3.1-kb *lacZ* fragment containing a minimal RBE (5'-GCGAGCGA-3'), the 2.7-kb cloned *BamHI* fragment RBE (5'-GAGCGAGCGAGC-3') from the ch-19 integration site, and a cloned CSF1 RBE (5'-GAGCGAGCGAGCGAGC-3') cDNA fragment (938 bp). Inclusion of a 2.1-kb fragment devoid of any RBE sequences was used as a negative control. Under these conditions, the three fragments with RBE sequences competed against labeled ch-19-Rep DNA to the same degree (Fig. 6). Nonspecific DNA did not show competition below a 100-fold molar excess, suggesting that the fragments containing non-ch-19 RBE sites were effective at binding Rep.

These competition results support the EM analysis and suggest that Rep68H6 is unable to discriminate between ch-19 and non-ch-19 RBE sites in the human genome. From all of these observations, our data support the idea that AAV Rep is binding to non-ch-19 RBE sequences in the human genome. If we utilize the in vitro binding results to the 8-bp RBE as a minimum viable site, then potentially 2×10^5 RBE sites exist per genome. This observation suggests that the amount of Rep expressed during a latent infection could be rate limiting for AAV targeted integration.

Detection of Rep in a latent infection. With approximately 2×10^5 RBE/genome, this raises the critical question as to how Rep can localize to the AAV ITR and the ch-19 target sequence for site-specific integration among numerous similar

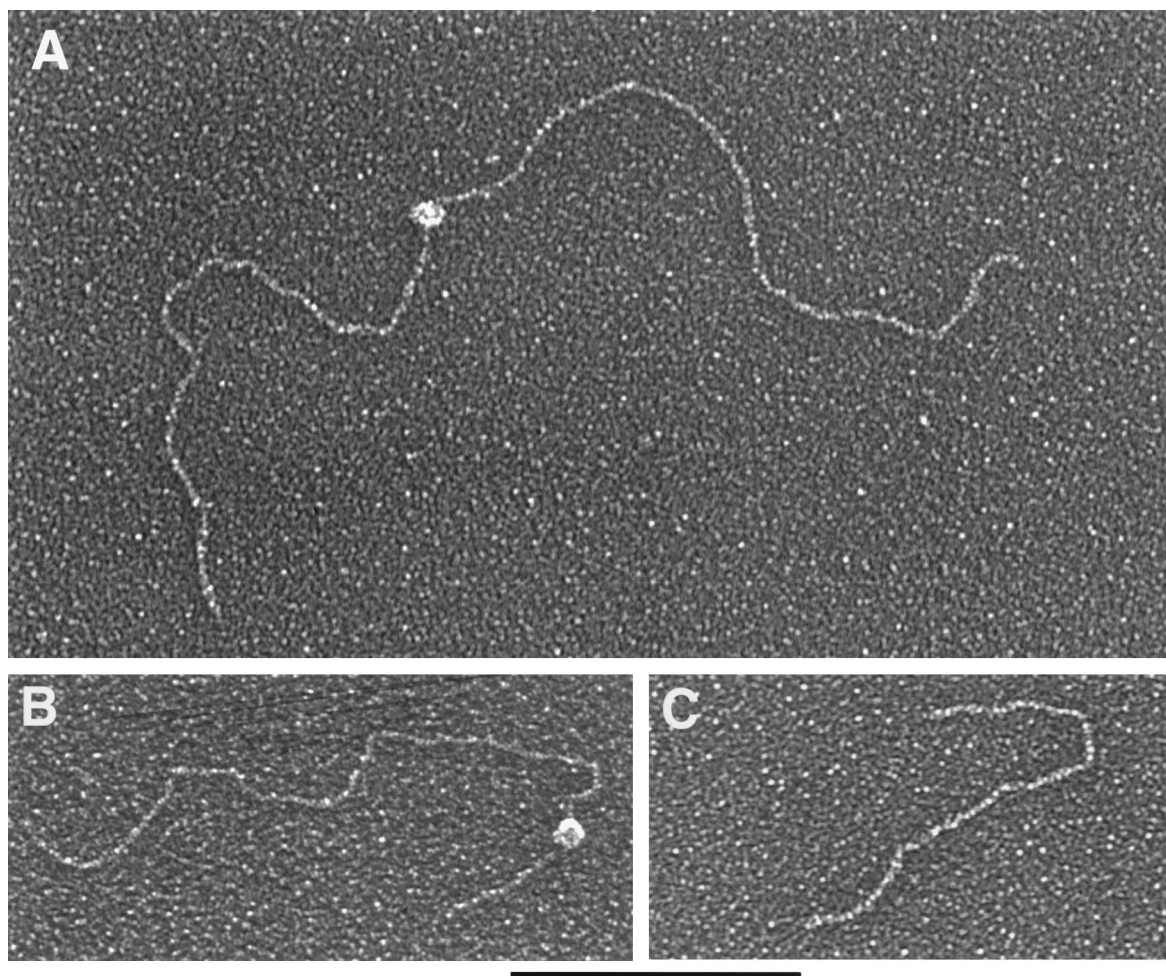


FIG. 4. EM analysis of Rep-DNA complexes at protein/DNA mass ratios of 1:1. (A) Rep68H6 protein bound to the center RBE of the 2.7-kb ch-19 integration fragment. (B) Rep68H6 protein bound to the asymmetric RBE of the 1.7-kb DNA fragment. (C) A 953-bp DNA fragment devoid of any RBE sequence and to which Rep68H6 did not bind. Protein binding and EM processing were carried out as described in Materials and Methods. Bar, 0.5 kb.

binding sites. Under this scenario, the amount of Rep expressed in a nonlytic infection becomes essential to its ability to carry out site-specific integration in the context of numerous RBE sequences. To determine the amount of Rep expressed in productive versus nonproductive infection, immunoprecipitation-Western analysis on 10^7 HeLa cells infected with Ad plus AAV or with AAV alone (MOI = 10), was carried out as described in Materials and Methods. Analysis of the Ad-AAV infection extract revealed that all four Rep proteins were produced in significant amounts (Fig. 7), as expected (3). In contrast to the Ad-AAV infection, only a faint band that corresponded to Rep52 was detected in the AAV-infected cells under similar exposures. Rep68 and Rep78 were detected in the AAV only infection after a 60-fold overexposure of the blot (Fig. 7). At this exposure, no protein in the mock-infected lane was detected, except for the rabbit immunoglobulin G that cross-reacted with the secondary antibody (Fig. 6). Using a serial dilution of purified Rep as a standard, the sensitivity of this assay was determined to be 200 pg (data not shown). Based on 10^7 cells, this value corresponds to a limit of detection of approximately 200 molecules of Rep78 or Rep68 per cell. After image quantification, we calculated approximately 1,000 to 4,000 Rep78 and Rep68 molecules per latent infected cell.

In addition to 1 to 4,000 molecules of Rep78 and Rep68, Rep52 (3,000 to 6,000 molecules) but not Rep40 was detected under these conditions.

AAV Rep induces ch-19 amplification. Most DNA-binding proteins associate with their specific sites at a rate that can exceed $10^9 \text{ M}^{-1} \text{ s}^{-1}$ (55). If we make a similar analogy to AAV Rep (using 200,000 sites per cell), this result suggests that there are sufficient Rep molecules available to interact with all potential RBEs within the human genome in a matter of minutes. Therefore, the Rep ability to target ch-19 must be related to subsequent steps after DNA binding. Previously, Linden et al. (30) determined that, in addition to the ch-19 RBE sequence, the targeting frequency was dependent on the presence of a *trs* site. This would imply that after Rep binds to a RBE site, the presence of a *trs* site is critical to targeting. To determine if Rep protein in HeLa cells influenced the ch-19 target sequence in the absence of ch-19 ITR targeting substrates, we cotransfected cells with Rep expressing plasmids and constructs carrying the neomycin resistance gene and isolated neomycin-specific clones. As shown in Fig. 8, we observed amplification and rearrangement of the ch-19-specific target DNA (Fig. 8, lanes 1, 2, and 3) in all cells that received Rep-expressing plasmids. Although we did not observe any targeted integra-

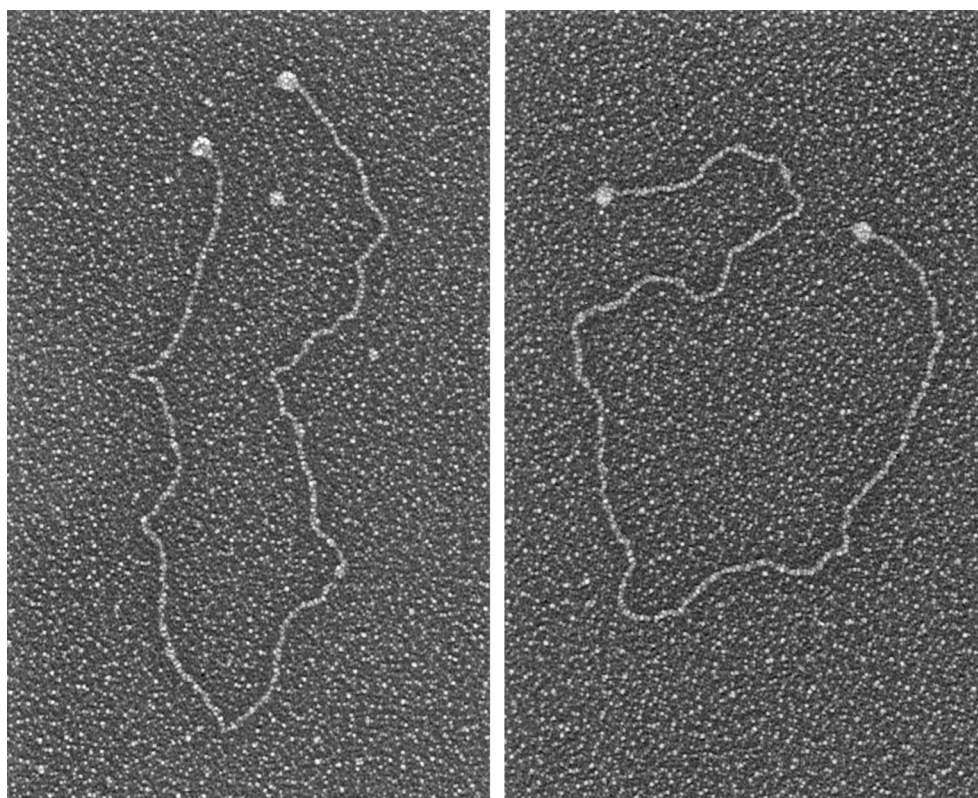


FIG. 5. EM analysis of Rep-DNA complexes at a protein/DNA mass ratio of 1:1. The DNA substrate was an AAV no-end viral sequence. Analysis was carried out as described in the text. All molecules displayed uniform Rep protein complexes associated with the terminal sequences (two examples are provided). Bar, 1 kb.

tion of the Neo plasmids into these cells (data not shown), these DNA rearrangements were similar to those described for targeted wt AAV integration. We never observed ch-19 amplification or rearrangement in cells that did not obtain Rep-expressing plasmids (data not shown). This observation supports *in vitro* studies demonstrating Rep-dependent replication of ch-19 sequences and suggests that even in the absence of functional AAV *cis*-acting targeting sequences, *trans*-acting Rep protein will act on the ch-19 integration locus.

Rep-dependent targeting to the mouse X chromosome. The above experiments support a role for Rep in AAV targeted integration by discriminating between ch-19 and non-ch-19 RBE sites through the presence of a *trs* sequence. In the presence of a *trs* site, Rep-induced DNA replication ensues potentially providing a substrate for targeted integration. The ch-19 RBE (8 bp), including the spacing and sequence of the *trs* (5 bp), most likely constitute a unique target sequence (15 bp) in the human genome, thereby ensuring AAV integration specificity. To test this hypothesis, we generated “knockin” mice that carried a portion of the ch-19 sequence on the mouse X chromosome (R. J. Samulski et al., manuscript submitted). Previously, we determined that the AAV target sequence was only conserved in human and nonhuman primates (R. J. Samulski, unpublished data), thereby providing a unique opportunity to study AAV targeted integration in this model. Primary mouse fibroblast cells were isolated from ch-19 mice and infected with AAV-Neo vectors carrying Rep coding sequences. Genomic DNA was digested with a no-cut enzyme for the vector sequence or the human ch-19 sequences and subjected to fractionation on agarose gel and characterized for targeted

integration by Southern analysis by using ch-19 and neomycin sequence-specific probes (Fig. 9). Using this assay, we determined the colocalization of the AAV Neo sequences with the human ch-19 sequences located on the mouse X chromosome. PCR junction analysis (data not shown) and filter-binding reactions confirmed the targeted integration of the AAV genome to this new chromosome location. These studies support the observations of Linden et al. (30) and Rizzuto et al. (40), which identified that ch-19 *cis* sequences are required for targeted integration when using an EBV shuttle vector in human cells or when integrated in the mouse and rat genome. In addition, these findings suggest that host enzymes required for AAV targeted integration are conserved in human and rodents and provide a mouse model for studying AAV site-specific integration.

DISCUSSION

In this study, we developed a Rep-dependent filter binding assay to evaluate the role of cellular RBE sites in AAV targeted integration. Using a highly purified source of Rep protein, we determined that human total genomic DNA can effectively compete for Rep binding. Southern analysis of genomic DNA before and after Rep-dependent filter binding demonstrated the specific retention of the ch-19 integration locus fragment containing the RBE and *trs* sequence, as well as non-ch-19 cellular DNA carrying only an RBE sequence. These results support the study by Wonderling et al. that identified alternative RBE sites existing in the human genome by BLAST analysis (58). Although we observed higher affinity for

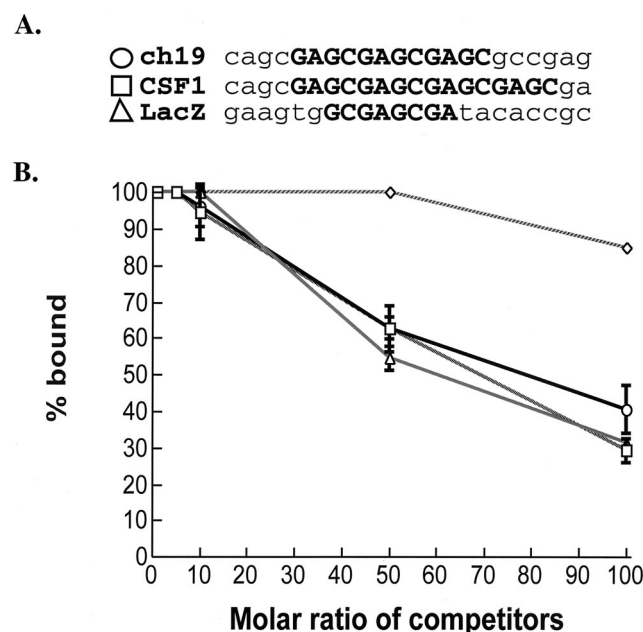


FIG. 6. Competition assay for Rep binding using different RBE DNA substrates. Filter-binding assays were assayed using 4 fmol of labeled *Bam*HI ch-19 fragment from plasmid pRE2 in a 50- μ l reaction volume. Rep68H6 and genomic DNA concentrations were the same as in the original assay conditions (see Materials and Methods). Increasing molar amounts of 3.1-kb *lacZ* fragment (Δ), 2.7-kb *Bam*HI ch-19 fragment (\circ), 938-bp CSF1 fragment (\square), and 2.1-kb nonspecific linearized pT7-7 BT plasmid (\diamond) were added to each reaction. Filter-binding reactions were carried out in a 96-well dot blot apparatus and washed with 2 ml of nonelution wash buffer before quantitation. Filters were then exposed to the phosphorimager cassette and the relative intensity was measured. Respective labeled dots were then cut and quantitated using a scintillation counter to corroborate the phosphorimager analysis. The data were then plotted as a function of fraction bound versus molar amount of competitor.

ch-19 specific probe

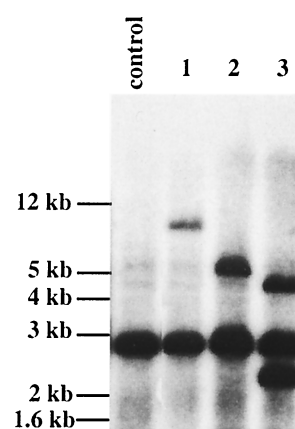


FIG. 8. Southern blot analysis of the amplified ch-19 region. HeLa cells were cotransfected with a Rep expression plasmid and Neo construct. Cells were G418 selected, and single colonies were picked and grown up. DNA from individual cell lines was digested with a restriction endonuclease that does not cut within the ch-19 region and was probed with a ch-19-specific probe. Control, untreated HeLa cells; lanes 1 to 3, cell lines derived after transfection with Rep expression plasmids.

ch-19 RBE sequences (2.7-fold) than for heterologous DNA carrying identical elements using this assay, this difference appears to be insufficient to explain AAV targeting. In fact, using these experimental conditions, we did not identify any features that would make the ch-19 locus a preferred site for Rep binding. For example, EM analysis of Rep-DNA complexes demonstrated highly uniform multimeric complexes on both ch-19 and non-ch-19 RBE sequences. These protein structures

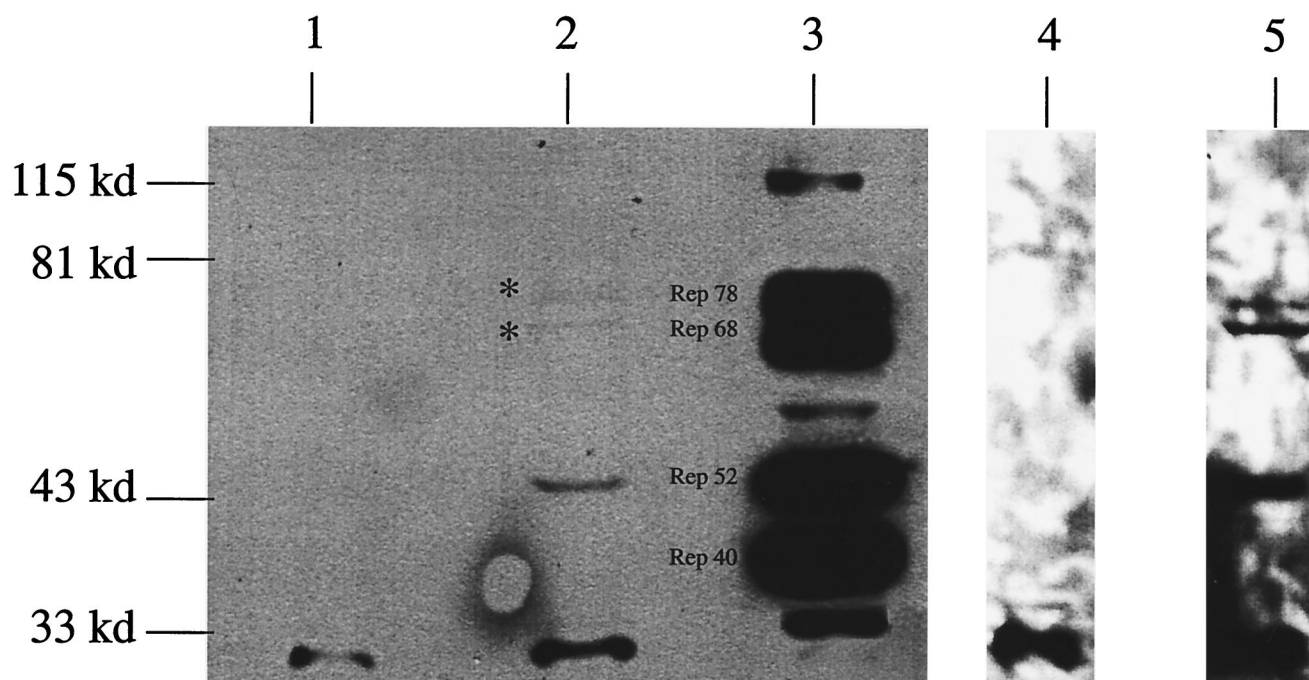


FIG. 7. Immunoprecipitation-Western analysis of Rep expression in AAV-infected cells. HeLa cells were infected with AAV alone or Ad-AAV, and protein extracts were isolated and assayed for Rep expression using AAV Rep-specific antibody. Lanes 1, 2, 4, and 5 show Western blot analysis of AAV Rep expression from 10^7 HeLa cells either mock infected (lanes 1 and 4) or AAV infected (lanes 2 and 5) that were exposed for 1 min (lanes 1 and 2) or 1 h (lanes 4 and 5). Lane 3 was a positive control of 5×10^6 Ad-AAV-coinfected HeLa cells demonstrating expression of all forms (78, 68, 52, and 40 kDa) of the AAV Rep proteins.

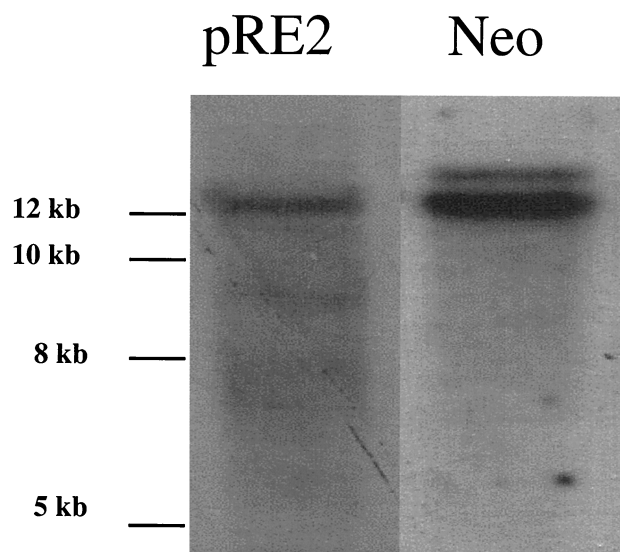


FIG. 9. Comigration of ch-19 target sequences from primary mouse cells with Rep-Neo vector DNA after analysis by the Southern blot technique. Primary mouse embryo fibroblast (PMEF) cells were infected with a recombinant Rep-Neo AAV virus. Latent cells which were Neo resistant were established as described in Materials and Methods. Lane 1, 10 μ g of *Xba*I-digested ch-19 (PMEF) DNA fractionated on agarose gel, transferred to membrane, and hybridized with a ch-19 sequence-specific probe; lane 2, 10 μ g of *Xba*I-digested ch-19 PMEF DNA was fractionated on agarose, transferred by Southern blotting, and hybridized with a Neo sequence-specific probe.

covered about 60 bp of DNA and appeared to be indistinguishable for ch-19, non-ch-19, and AAV ITR RBE sites. We never observed Rep binding to control DNA substrates lacking the RBE site. In fact, under these conditions we were able to establish Rep-binding complexes with minimum RBE elements of 8 bp, suggesting that as many as 200,000 Rep-binding sites may exist in the human genome. These results imply that Rep may interact with numerous RBE sites distributed across the human genome as part of the integration mechanism.

Rep levels in a latent infection. Based on this result, the amount of Rep in a latent infection could be considered as a rate-limiting step for AAV site-specific integration. For this reason, we determined the amount of Rep expressed in the first 24 h of a nonlytic infection. Our analysis suggested that 1 to 4,000 copies of Rep78 and Rep68 are available for facilitating targeted integration. At present, the rate association constant for Rep and cellular RBE sites have not been determined. Other DNA-binding proteins (e.g., *Drosophila* doublesex and Lambda Cro) typically have rate association constants on the order of 10^6 (6) and 10^8 (22), respectively. Extrapolating from other known DNA association constants and the time observed for AAV integration (24 h) (34), this would suggest that the level of Rep78 and Rep68 present in a nonproductive infection (1 to 4,000 molecules) should be sufficient to interact with all 200,000 potential RBE sites. While it is tempting to suggest this argument, determination of Rep-RBE association constants need to be established before firm conclusions can be made. However, even if Rep is inefficient at binding, it is important to note that, in vivo, it is unlikely that all 200,000 sites are accessible to Rep binding due to their chromatin structure. In fact, a recent study has shown that the ch-19 site is DNase hypersensitive (27), implying a chromatin structure that is accessible for protein interactions. This observation implies that ch-19 is always available for Rep interaction, whereas other sites may be subject to local chromatin

environment during the cell cycle. These factors strengthen the notion that 1,000 to 4,000 molecules of Rep could be sufficient for targeting the viral genome in the absence of any other features.

Rep protein complexes. Our characterization of purified Rep-DNA complexes by EM analysis represents the first visual identification of a multimeric-protein complex interacting with the Rep binding site. The fact that Rep complexes were multimeric is not surprising since it has been reported that Rep forms multiple protein-DNA complexes (33) and may potentially bind as a hexamer (48). The observation that Rep can span a region of DNA of about 60 bp, however, does provide an explanation as to why Rep does not bind efficiently to small oligonucleotides as described by Chiorini et al. (4). It is possible that the RBE DNA induces multimerization of the Rep68 protein since the Rep protein concentration used in the EM experiments (16 nM) was significantly lower than that for Rep when it exists as a monomer in solution (380 nM or 25 ng/ μ l). It is interesting to note that previous studies have suggested Rep-Rep interactions when there is binding to the DNA RBE motif (32, 33). Our observations clearly establish that Rep complexes interacting with either AAV ITR, ch-19 RBE, or minimum RBE element derived from plasmid DNA form identical protein structures (Fig. 4). It was interesting to note that, given that the 953-bp fragment with no RBE showed no nonspecific binding, the RBE-containing units had 11 and 15% nonspecific binding (Table 1). This observation suggests that the presence of a strong RBE increases the local concentration of Rep enough to allow interaction with low-affinity sites on the same DNA strand. This may be a factor that is important in vivo, where competition for Rep (RBE sites) and the initiation for site-specific recombination may be influenced by the local concentration of Rep molecules (Rep-Rep interactions). However, further experiments are required to evaluate this hypothesis.

Regardless, as previously indicated by gel shift assays, EM analysis also revealed a higher affinity of Rep for the ITR (Fig. 5) when it was compared to either ch-19 or analogous RBE sequences (Table 1). Based on this observation we hypothesize that Rep levels in a newly infected cell will allow constant occupation of the AAV ITR sequences compared to transient binding to cellular RBE elements other than ch-19. This would imply that a preformed AAV ITR-Rep complex is a separate substrate that may interact with chromosomal RBE sites or chromosomal Rep-RBE complexes. If the ITR-Rep complex interacts with a Rep-chromosome complex as a separate entity, then the formation of the appropriate Rep-chromosome RBE complex would become the rate-limiting determinant for targeted recombination.

Cellular factors. It is important to take into account that all of our binding analyses were performed with purified Rep only. Recent evidence has demonstrated a role for high-mobility group (HMG) proteins in NS1 nicking of the MVM genome (8). This protein has also enhanced Rep binding and nicking activity on the AAV ITR (7) and has enhanced in vitro targeting of AAV substrates (9). HMG proteins have been implicated in bending DNA and making it more flexible (36, 38, 39). Therefore, our observations with purified Rep would likely be enhanced in the presence of HMG. Cellular proteins could clearly impact the efficiency of Rep-mediated integration (62), and this fact points to a limitation of our study design. The in vitro observations we have described, however, were extended by using normal diploid cells derived from a ch-19 mouse (see below).

Rep-dependent ch-19 replication. Our experiments that assayed for Rep effect on ch-19 yielded interesting observations

with regard to AAV targeted integration. The ability to amplify ch-19 in the absence of viral integration substrates strongly suggest that Rep-mediated replication is a primary step for AAV targeting. These observations support earlier studies by Urcelay et al. (54) that described an *in vitro* Rep-dependent replication of ch-19 DNA carrying the RBE-*trs* site. From these observations, it appears that a critical step in the AAV integration process involves Rep-dependent nicking of the ch-19 substrate. Our studies also demonstrate cellular amplification in the absence of viral targeting sequences, suggesting that Rep initiates a replication event on ch-19 that results in amplification and rearrangement. This amplification and rearrangement may result from Rep-dependent unscheduled initiation of ch-19 replication. These observations also suggest the possibility that the head-to-tail configuration of viral integrants appears as a by-product of this replication event. It is interesting to note that proviral structures for AAV vectors devoid of Rep result in identical head-to-tail concatemers, albeit at random sites in the genome (25, 34, 62). In addition, recent analysis of simian virus 40 integration has documented identical head-to-tail proviral structures, implying that this may be a universal cellular mechanism for "amplified" integration (52). These data imply that host machinery is responsible for the head-to-tail amplification and that the role of Rep is to direct the recombination event through initiation of replication on a virus-like origin (RBE-*trs*) located on ch-19. These observations also provide an explanation as to why latent proviral structures (head to tail) generated by host enzymes do not resemble the Rep-dependent viral replication intermediates (head to head) seen in a lytic infection. In addition, EM data suggest that the Rep complex may be associated with the viral TR sequences prior to initiating replication on ch-19. It is still undetermined whether the viral Rep-TR complex requires nicking in order to recombine with the ch-19 target sequence. While rescue of proviral AAV genomes requires functional terminal repeats for replication and packaging, the precise role in integration is still undefined. These data suggest that if Rep independently forms a complex on ch-19 and initiates replication, this may be a hot spot for Rep-Rep DNA complexes to assemble, in a manner similar to that of the punctate replication centers described by Hunter et al. (16), for wt AAV lytic infection.

Animal model for AAV targeting. *In vivo* characterization of Rep-dependent replication of ch-19 suggests that AAV targeted integration is dependent on a replication and not on a viral integrase-like mechanism for targeted integration. More importantly, these observations, coupled with earlier studies identifying the ch-19 minimum *cis* sequences, suggest that AAV targeted integration can be characterized both *in vitro* and *in vivo*. This may be important in determining the targeted integration potential of different cell types. In an effort to generate an animal model suitable for AAV targeted integration, we utilized a *cis* sequence (2.7 kb) from ch-19 to generate a knockin transgenic mouse (Samulski et al., submitted). This animal now carries the targeting sequences on the X chromosome located in the HPRT site. Our initial characterization using Rep-Neo vectors demonstrated targeted integration into the ch-19 sequences now located on the mouse X chromosome. These observations confirm two points: (i) host enzymes involved in AAV targeted integration are conserved between rodents and humans, and (ii) the targeting locus carries the sequence information (e.g., chromatin structure) required to identify this region for recombination. This premise is supported by the fact that Lamartina et al. determined that this sequence carries a DNase hypersensitive site which may facilitate the targeting process (27). Recently, Rizzuto et al. (40)

established a rat model carrying a different but overlapping portion of ch-19 for targeting. From these two experimental animal models, a minimum sequence of 1.6 kb was found to be common, suggesting that this sequence may be sufficient for targeting *in vivo*. Prior to these studies, Linden et al. (29, 30) established by using an integration assay dependent on an EBV episome carrying the ch-19 sequence that a minimum of 33 bp was sufficient for targeting. Previously, a sequence 5' to the RBE was determined to cause instability in the EBV system. This is the same region identified by Lamartina et al. for DNase hypersensitivity. It is interesting to speculate that these sequences (which are not required in the episomal system) may be required for targeting to the chromosomal locus. It remains to be determined if the 33-bp sequence alone will retain targeting in a chromosomal locus. Regardless, the studies involving these animals strongly support the idea that this region of human ch-19 containing the RBE-*trs* sequences is unique in the human and simian genome and will function when moved to a new location. This animal model should be useful now for studying AAV integration *in vivo* in diploid cells. More importantly, this model may provide a research tool for studying chromosomal targeting using AAV plasmid vectors.

At present, 1% of AAV infecting viral genomes integrate with approximately 70 to 90% of these proviruses targeted to ch-19 (35). Though this is an efficient reaction for targeting, the overall integration frequency is marginal at best. We can identify at least five successive steps that would impact AAV targeted integration after infection: (i) after successful viral entry and uncoating, the conversion of single-stranded to double-stranded DNA, providing a template competent for mRNA expression; (ii) expression of Rep proteins; (iii) Rep interaction and complex formation with viral ITR and/or chromosomal RBEs; (iv) Rep-dependent replication of the ch-19 locus; and (v) Rep-ITR complexes interacting with Rep-ch-19 replicated substrates, facilitating targeted recombination and resulting in head-to-tail proviral structures via host enzymes.

In conclusion, we have utilized a highly purified source of Rep to investigate the role of this protein in AAV targeted integration. It is apparent that Rep can interact with numerous RBE like sequences in the human genome (ca. $>10^5$ sites) and that the differential affinity of Rep for these sites is unlikely to provide the specificity of targeted integration. EM studies suggest that large multimeric Rep protein complex binds to degenerate RBE sites as well as to the ch-19 site. Rep initiates replications at the ch-19 target locus in the presence or absence of a viral genome. This event leads to amplification of ch-19 that may continue after recombination of the viral genome to generate concatemeric proviral structures. Finally, these recombination products can be generated at both endogenous ch-19 position, as well as at heterologous sites generated in a mouse model. Further analysis of these observations should provide a molecular understanding for the role of Rep in AAV site-specific recombination *in vivo*.

ACKNOWLEDGMENTS

We thank Michael Topal and Terry Van Dyke for constructive comments and criticisms of this study and the manuscript. We also thank Jack Griffith for providing help and expertise with EM analysis and Irene Zolotukhin and N. Muzyczka for protein purification training.

This work was supported by grants DK51880 and HL048347 to R.J.S., 2T32A107419 to S.M.Y., and CA70343 and GM31819 to J.G. for the support of N.D.

REFERENCES

1. Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. 1998. Current protocols in molecular biology, vol. 1.

- John Wiley & Sons, Inc., New York, N.Y.
2. Balague, C., M. Kalla, and W. Zhang. 1997. Adeno-associated virus Rep78 protein and terminal repeats enhance integration of DNA sequences into the cellular genome. *J. Virol.* **71**:3299–3306.
 3. Berns, K. I. 1996. Parvoviridae: the viruses and their replication, p. 2173–2197. *In* B. N. Fields (ed.), *Fields virology*, 3rd ed., vol. 2. Raven, Philadelphia, Pa.
 4. Chiorini, J. A., S. M. Wiener, R. A. Owens, S. R. M. Kyostio, R. M. Kotin, and B. Safer. 1994. Sequence requirements for stable binding and function of Rep68 on the adeno-associated virus type 2 inverted terminal repeats. *J. Virol.* **68**:7448–7457.
 5. Chiorini, J. A., L. Yang, B. Safer, and R. M. Kotin. 1995. Determination of adeno-associated virus Rep68 and Rep78 binding site by random sequence oligonucleotide selection. *J. Virol.* **69**:7334–7338.
 6. Cho, S., and P. Wensink. 1997. DNA binding by the male and female doublesex proteins of *Drosophila melanogaster*. *J. Biol. Chem.* **272**:3185–3189.
 7. Costello, E., P. Saudan, E. Winocour, L. Pizer, and P. Beard. 1997. High mobility group chromosomal protein 1 binds to the adeno-associated virus replication protein (Rep) and promotes Rep-mediated site-specific cleavage of DNA, ATPase activity and transcriptional repression. *EMBO J.* **16**:5943–5954.
 8. Cotmore, S. F., and P. Tattersall. 1998. High-mobility group 1/2 proteins are essential for initiating rolling-circle-type DNA replication at a parvovirus hairpin origin. *J. Virol.* **72**:8477–8484.
 9. Dyall, J., P. Szabo, and K. I. Berns. 1999. Adeno-associated virus (AAV) site-specific integration: formation of AAV-AAVS1 junctions in an in vitro system. *Proc. Natl. Acad. Sci. USA* **96**:12849–12854.
 10. Ferrari, F. K., T. Samulski, T. Shenk, and R. J. Samulski. 1996. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J. Virol.* **70**:3227–3234.
 11. Fuller, R. S., B. E. Funnell, and A. Kornberg. 1984. The DnaA protein complex with the *E. coli* chromosomal replication origin (oriC) and other DNA sites. *Cell* **38**:889–900.
 12. Gavin, D. K., S. M. Young, Jr., W. Xiao, B. Temple, C. R. Abernathy, D. J. Pereira, N. Muzyczka, and R. J. Samulski. 1999. Charge-to-alanine mutagenesis of the adeno-associated virus type 2 Rep78/68 proteins yields temperature-sensitive and magnesium-dependent variants. *J. Virol.* **73**:9433–9445.
 13. Gottesman, S., E. Halpern, and P. Trisler. 1981. Role of *sulA* and *sulB* in filamentation by *lon* mutants of *Escherichia coli* K-12. *J. Bacteriol.* **148**:265–273.
 14. Griffith, J., and G. Christensen. 1978. Electron microscope visualization of chromatin and other DNA-protein complexes. *Annu. Rev. Biophys. Bioeng.* **7**:19–35.
 15. Hermonat, P. L., M. A. Labow, R. Wright, K. I. Berns, and N. Muzyczka. 1984. Genetics of adeno-associated virus: isolation and preliminary characterization of adeno-associated virus type 2 mutants. *J. Virol.* **51**:329–333.
 16. Hunter, L. A., and R. J. Samulski. 1992. Colocalization of adeno-associated virus Rep and capsid proteins in the nuclei of infected cells. *J. Virol.* **66**:317–324.
 17. Im, D.-S., and N. Muzyczka. 1990. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. *Cell* **61**:447–457.
 18. Im, D.-S., and N. Muzyczka. 1989. Factors that bind to the AAV terminal repeats. *J. Virol.* **63**:3095–3104.
 19. Im, D. S., and N. Muzyczka. 1992. Partial purification of adeno-associated virus Rep78, Rep52, and Rep40 and their biochemical characterization. *J. Virol.* **66**:1119–1128.
 20. Kawasaki, E. S., M. B. Ladner, A. M. Wang, J. Van Arsdel, M. K. Warren, M. Y. Coyne, V. L. Schweickart, M. T. Lee, K. J. Wilson, A. Boosman, E. R. Stanley, P. Ralph, and D. F. Mark. 1985. Molecular cloning of a complementary DNA encoding human macrophage-specific colony-stimulating factor (CSF-1). *Science* **230**:291–296.
 21. Kearns, W. G., S. A. Afione, S. B. Fulmer, M. G. Pang, D. Erikson, M. Egan, M. J. Landrum, T. R. Flotte, and G. R. Cutting. 1996. Recombinant adeno-associated virus (AAV-CFTR) vectors do not integrate in a site-specific fashion in an immortalized epithelial cell line. *Gene Ther.* **3**:748–755.
 22. Kim, J., Y. Takeda, B. Matthews, and W. Anderson. 1987. Kinetic studies on Cro repressor-operator DNA interaction. *J. Mol. Biol.* **196**:149–158.
 23. Kotin, R. M., R. M. Linden, and K. I. Berns. 1992. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J.* **11**:5071–5078.
 24. Kotin, R. M., J. C. Menninger, D. C. Ward, and K. I. Berns. 1991. Mapping and direct visualization of a region-specific viral DNA integration site on chromosome 19q13-qter. *Genomics* **10**:831–834.
 25. Kotin, R. M., M. Siniscalco, R. J. Samulski, X. D. Zhu, L. Hunter, C. A. Laughlin, S. McLaughlin, N. Muzyczka, M. Rocchi, and K. I. Berns. 1990. Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. USA* **87**:2211–2215.
 26. Ladner, M. B., G. A. Martin, J. A. Noble, D. M. Nikoloff, R. Tal, E. S. Kawasaki, and T. J. White. 1987. Human CSF-1: gene structure and alternative splicing of mRNA precursors. *EMBO J.* **6**:2694–2698.
 27. Lamartina, S., E. Sporeno, and C. Toniatti. 1999. In vitro and in vivo analysis of the AAV integration site in human chromosome 19. *J. Gene Med.* **1**:68.
 28. Lee, S., L. Cavallo, and J. Griffith. 1997. Human p53 binds Holliday junctions strongly and facilitates their cleavage. *J. Biol. Chem.* **272**:7532–7539.
 29. Linden, R. M., P. Ward, C. Giraud, E. Winocour, and K. Berns. 1996. Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. USA* **93**:11288–11294.
 30. Linden, R. M., E. Winocour, and K. I. Berns. 1996. The recombination signals for adeno-associated virus site-specific integration. *Proc. Natl. Acad. Sci. USA* **93**:7966–7972.
 31. McCarty, D. M., T. H. Ni, and N. Muzyczka. 1992. Analysis of mutations in adeno-associated virus Rep protein in vivo and in vitro. *J. Virol.* **66**:4050–4057.
 32. McCarty, D. M., D. J. Pereira, I. Zolotukhin, X. Zhou, J. H. Ryan, and N. Muzyczka. 1994. Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. *J. Virol.* **68**:4988–4997.
 33. McCarty, D. M., J. H. Ryan, S. Zolotukhin, X. Zhou, and N. Muzyczka. 1994. Interaction of the adeno-associated virus Rep protein with a sequence within the A palindrome of the viral terminal repeat. *J. Virol.* **68**:4998–5006.
 34. McLaughlin, S. K., P. Collis, P. L. Hermonat, and N. Muzyczka. 1988. Adeno-associated virus general transduction vectors: analysis of proviral structures. *J. Virol.* **62**:1963–1973.
 35. Muzyczka, N. 1992. Use of adeno-associated virus as a general transduction vector for mammalian cells. *Curr. Top. Microbiol. Immunol.* **158**:97–129.
 36. Onate, S. A., P. Prendergast, J. P. Wagner, M. Nissen, R. Reeves, D. E. Pettijohn, and D. P. Edwards. 1994. The DNA-binding protein HMG-1 enhances progesterone receptor binding to its target DNA sequences. *Mol. Cell. Biol.* **14**:3376–3391.
 37. Owens, R. A., M. D. Weitzman, S. R. M. Kyostio, and B. J. Carter. 1993. Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. *J. Virol.* **67**:997–1005.
 38. Paull, T. T., M. J. Haykinson, and R. C. Johnson. 1993. The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. *Genes Dev.* **7**:1521–1534.
 39. Pil, P. M., C. S. Chow, and S. J. Lippard. 1993. High-mobility-group 1 protein mediates DNA bending as determined by ring closures. *Proc. Natl. Acad. Sci. USA* **90**:9465–9469.
 40. Rizzuto, G., B. Gorgoni, M. Capelletti, D. Lazzaro, I. Gloaguen, V. Poli, A. Sgura, D. Cimini, G. Ciliberto, R. Cortese, E. Fattori, and N. La Monica. 1999. Development of animal models for adeno-associated virus site-specific integration. *J. Virol.* **73**:2517–2526.
 41. Rutledge, R. A., and D. W. Russell. 1997. Adeno-associated virus vector integration junctions. *J. Virol.* **71**:8429–8436.
 42. Ryan, J. H., S. Zolotukhin, and N. Muzyczka. 1996. Sequence requirements for binding of Rep68 to the adeno-associated virus terminal repeats. *J. Virol.* **70**:1542–1553.
 43. Samulski, R. J. 1993. Adeno-associated virus: integration at a specific chromosomal locus. *Curr. Opin. Genet. Dev.* **3**:74–80.
 44. Samulski, R. J., L.-S. Chang, and T. Shenk. 1989. Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. *J. Virol.* **63**:3822–3828.
 45. Samulski, R. J., M. Sally, and N. Muzyczka. 1999. Adeno-associated viral vectors, p. 131–172. *In* The development of human gene therapy. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 46. Samulski, R. J., X. Zhu, X. Xiao, J. D. Brook, D. E. Housman, N. Epstein, and L. A. Hunter. 1991. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J.* **10**:3941–3950. (Erratum, **11**:1228, 1992.)
 47. Shelling, A. N., and M. G. Smith. 1994. Targeted integration of transduced and infected adeno-associated virus vectors containing the neomycin resistance gene. *Gene Ther.* **1**:165–169.
 48. Smith, R. H., A. J. Spano, and R. M. Kotin. 1997. The Rep78 gene product of adeno-associated virus (AAV) self-associates to form a hexameric complex in the presence of AAV *ori* sequences. *J. Virol.* **71**:4461–4471.
 49. Snyder, R. O., R. J. Samulski, and N. Muzyczka. 1990. In vitro resolution of covalently joined AAV chromosome ends. *Cell* **60**:105–133.
 50. Srivastava, A., E. W. Lusby, and K. I. Berns. 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J. Virol.* **45**:555–564.
 51. Surosky, R. T., M. Urabe, S. G. Godwin, S. A. McQuiston, G. J. Kurtzman, K. Ozawa, and G. Natsoulis. 1997. Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. *J. Virol.* **71**:7951–7959.
 52. Syu, L., and M. Fluck. 1997. Site-specific in situ amplification of the integrated polyomavirus genome: a case for a context-specific over-replication model of gene amplification. *J. Mol. Biol.* **271**:76–99.
 53. Tratschin, J.-D., I. L. Miller, and B. J. Carter. 1984. Genetic analysis of adeno-associated virus: properties of deletion mutants constructed in vitro and evidence for an adeno-associated virus replication function. *J. Virol.* **51**:611–619.
 54. Urclay, E., P. Ward, S. M. Wiener, B. Safer, and R. M. Kotin. 1995. Asymmetric replication in vitro from a human sequence element is dependent

- dent on adeno-associated virus Rep protein. *J. Virol.* **69**:2038–2046.
55. von Hippel, P., and O. Berg. 1989. Facilitated target location in biological systems. *J. Biol. Chem.* **264**:675–678.
 56. Walsh, C. E., J. M. Liu, N. S. Young, A. W. Nienhuis, and R. J. Samulski. 1992. Regulated high level expression of a human γ -globin gene introduced into erythroid cells by adeno-associated virus vector. *Proc. Natl. Acad. Sci. USA* **89**:7257–7261.
 57. Weitzman, M. D., S. R. M. Kyostio, R. M. Kotin, and R. A. Owens. 1994. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl. Acad. Sci. USA* **91**:5808–5812.
 58. Wonderling, R., and R. Owens. 1997. Binding sites for adeno-associated virus Rep proteins within the human genome. *J. Virol.* **71**:2528–2534.
 59. Wonderling, R. S., S. R. M. Kyostio, and R. A. Owens. 1995. A maltose-binding protein/adeno-associated virus Rep68 fusion protein has DNA-RNA helicase and ATPase activities. *J. Virol.* **69**:3542–3548.
 60. Xiao, W. 1996. Ph.D. thesis. University of North Carolina at Chapel Hill, Chapel Hill.
 61. Xiao, X., W. Xiao, J. Li, and R. J. Samulski. 1997. A novel 165-base-pair terminal repeat sequence is the sole *cis* requirement for the adeno-associated virus life cycle. *J. Virol.* **71**:941–948.
 62. Yang, C. C., X. Xiao, X. Zhu, D. C. Ansardi, N. D. Epstein, M. R. Frey, A. G. Matera, and R. J. Samulski. 1997. Cellular recombination pathways and viral terminal repeat hairpin structures are sufficient for adeno-associated virus integration in vivo and in vitro. *J. Virol.* **71**:9231–9247.
 - 62a. Young, S. M., Jr. 2000. Ph.D. thesis. University of North Carolina at Chapel Hill, Chapel Hill.
 63. Zhu, X. 1993. Ph.D. thesis. University of Pittsburgh, Pittsburgh, Pa.